Сравнение методов детектирования трехмерных объектов в задаче автономного вождения транспортных средств

А.А. Агафонов[1], А.С. Юмаганов[1]

[1]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Задача управлением движением автономных транспортных средств является одной из наиболее актуальных задач как с исследовательской, так и с практической точки зрения. Для ее решения необходимо рассмотреть целый ряд подзадач, включая локализацию и маппинг, обнаружение статических и динамических объектов, планирование маневров, контроль и т.д. В данной работе рассматривается одна из важнейших задач системы управления беспилотным автомобилем: задача обнаружения и локализации трехмерных объектов. Исследуются существующие методы детектирования, использующие информацию с различных датчиков транспортного средства: камер и LIDAR датчиков. Представлены результаты экспериментальных исследований качества работы данных методов на реальных и синтетических данных, полученных с использованием CARLA – программного обеспечения для моделирования и исследования автономного вождения .

1. Введение

В настоящее время задача создания автономных транспортных средств является одной из наиболее актуальных проблем в транспортной области. Такие транспортные средства призваны снизить и оптимизировать загруженность дорог, минимизировать количество дорожно-транспортных происшествий, обеспечить возможность перевозки грузов в опасных для человека условиях, расширить возможности использования автомобилей для людей с ограниченными физическими возможностями и т.д. Актуальность задачи подчеркивается как наличием большого числа исследований по данной тематике, так и работой большого числа автомобильных концернов над этой проблемой.

Архитектура системы управления беспилотных автомобилей обычно состоит из двух основных частей: системы восприятия и системы принятия решений [1]. Система восприятия отвечает за оценку состояния автомобиля и за создание внутреннего представления об окружающей среде, используя данные, полученные с помощью бортовых датчиков (таких как LIDAR, радар, камера, GPS, инерциальный измерительный блок, одометр), а также предварительную информацию о дорожной сети, правилах дорожного движения, динамике автомобиля и т.д.

Система восприятия, как правило, делится на несколько подсистем, отвечающих за различные задачи, например, локализация беспилотного автомобиля, обнаружение статических препятствий, обнаружение и отслеживание движущихся объектов, обнаружение и распознавание знаков дорожного движения и т.д.

Система принятия решений отвечает за движение автомобиля от его исходного положения до конечной цели, определенной пользователем, с учетом текущего состояния автомобиля, внутреннего представления окружающей среды, правил дорожного движения, а также безопасности и комфорта пассажиров. Система принятия решений в свою очередь разделяется на подсистемы, отвечающие за такие задачи, как планирование маршрута, планирование траектории движения, анализ дорожной ситуации, обход препятствий и контроль движения. Однако такое разделение на подсистемы несколько размыто, и в литературе встречаются различные вариации такого разделения [1].

В данной работе рассматриваются методы решения одной из важнейших задач системы восприятия беспилотного автомобиля – обнаружение и локализации трехмерных объектов. Методы обнаружения и локализации трехмерных объектов условно можно разделить на три группы по используемым датчикам: методы на основе камеры, на основе LIDAR датчика, на основе камеры и датчика LIDAR.

Известно множество методов решения названной выше задачи, использующих данные с камеры. Авторы [2,3] для обнаружения трехмерных контуров транспортных средств использовали шаблоны 3D моделей различных видов этих транспортных средств (например, для автомобилей: седан, хэтчбек, внедорожник и т.д.). В работах [4,5] авторами представлены методы обнаружения, которые для построения трехмерных контуров транспортных средств используют информацию о геометрической связи между двумерным и трехмерным контуром объекта. Эти методы используют данные, полученные с помощью обычной фотокамеры, средняя стоимость которой существенно ниже, чем стоимость LIDAR датчиков. Однако качество получаемых данными методами результатов обнаружения существенно уступает результатам методов, использующих информацию, полученную с помощью LIDAR датчиков. Это обусловлено тем, что снимки, полученные с помощью обычной камеры, не содержат информации о расстоянии до запечатленных объектов.

Широко распространение получили методы обнаружения трехмерных объектов, использующих полученные с помощью датчика LIDAR облака точек (points cloud). В работах [6,7] представлены методы обнаружения, в котором облако точек разбивается на воксели и для каждого вокселя различными способами формируется векторное описание. Затем используется сеть для предложения регионов (Region Proposal Network (RPN)), с помощью которой получаются трехмерные контуры объектов. В работе [8] векторное представление облака точек формируется путем его проекции на двумерное изображение, которое представляет собой вид сверху (bird’s-eye view (BEV)).

Высокое качество обнаружения демонстрируют также методы, использующие информацию с двух различных датчиков – LIDAR и монокулярная камера. В работе [9] из облака точек формируется две проекции – вид сверху и фронтальная проекция, которые совместно с фронтальным изображением монокулярной камеры являются входными данными данного метода. С помощью RPN сети для проекции «вид сверху» генерируются предположения о расположении объектов (3D object proposals), которые затем проецируются на две другие проекции. Для объединения полученных признаков и получения окончательного результата используется глубокая нейронная сеть. Авторы метода AVOD [10] улучшили качество работы метода [9] для небольших объектов, путем генерации предположений о расположении объектов не только для проекции «вид сверху», но и для RGB изображения с камеры.

В процессе разработки системы управления беспилотным автомобилем для испытания работы ее отдельных систем или всей системы в целом часто применяются средства симуляции. Это позволяет сэкономить время и финансовые ресурсы. Кроме того, применение симуляторов позволяет эмулировать различные дорожные ситуации, в том числе и опасные для жизни и здоровья людей, при проведении испытаний той или иной системы беспилотного автомобиля. Однако, настроенная или обученная только на синтетических данных система может показывать существенно отличающиеся результаты при ее проверке на реальных данных.

В данной работе проводится испытание известных методов обнаружения и локализации трехмерных объектов на данных, полученных из среды симуляции с открытым исходным кодом CARLA [11], и реальных данных, полученных из открытой базы KITTI[12].

2. Методы обнаружения и локализации трехмерных объектов

Для анализа были выбраны два метода обнаружения [7,13], использующие данные LIDAR датчика, и метод [10], использующий информацию с двух датчиков – LIDAR и монокулярной камеры. Рассмотрим подробнее каждый из данных методов.

2.1. AVOD

Общая схема метода AVOD (Aggregate View Object Detection) представлена на рисунке 1.

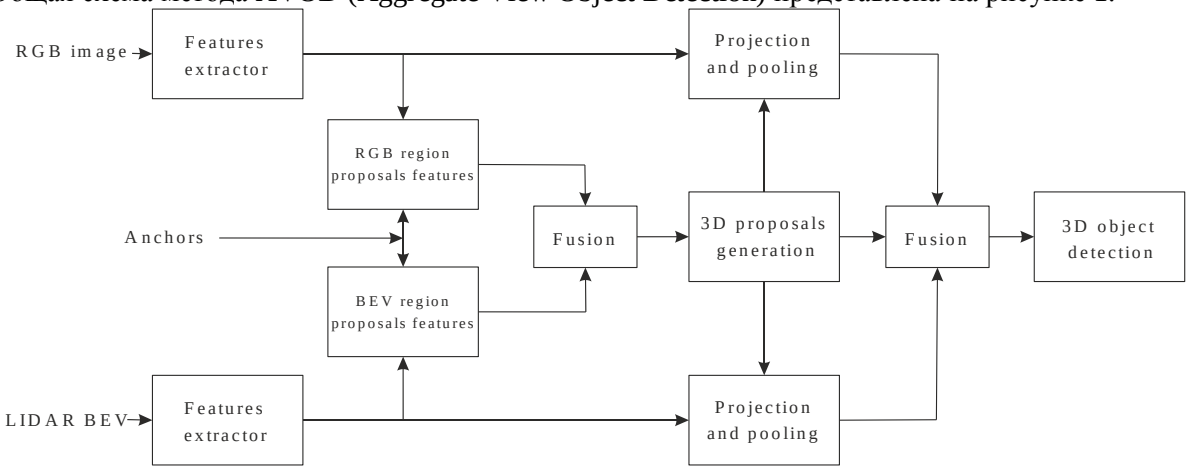

Рисунок 1. Структурная схема архитектуры метода AVOD.

Входными данными для данной сети являются изображение, полученное с помощью монокулярной камеры, и проекция облака точек «вид сверху» (BEV), полученного с помощью LIDAR датчиков. Для извлечения признаков из входных данных используются две идентичные encoder-decoder сети, в основе которых лежит модифицированная сверточная нейронная сеть VGG16[14]. Полученные векторы признаков используются в дальнейшем RPN сетью и сетью детектирования.

RPN сеть вычисляет разницу между заранее сформированными трехмерными контурами и истинными контурами объектов. Эти предварительно сгенерированные трехмерные контуры называются якорями (anchors). Для каждого якоря формируется два набора векторов признаков одинакового размера, соответствующих фронтальному изображению с камеры и BEV проекции, которые объединяются в один вектор с помощью операции поэлементного усреднения. Затем RPN сеть, представляющая собой полносвязную нейронную сеть, используя полученные векторные описания якорей, вычисляет разницу в положении и размере текущего предполагаемого региона от истинного положения и размера объектов и дает оценку того, содержит ли анализируемый контур «объект» или он является фоном. Фоновые якоря определяются путем вычисления метрики IoU (Intersection over Union) на BEV проекции между якорем и истинными контурами объектов. Для объектов класса «автомобиль» якоря со значением метрики IoU меньше 0.3 отмечаются как фоновые, а якоря со значением данной метрики больше 0.5 – как содержащие объект. Для фильтрации полученных регионов (якорей) используется метод подавления немаксимумов (non-maximum suppression) с пороговом значением метрики IoU равным 0.8 на BEV проекции.

В результате на вход сети детектирования поступает 1024 лучших по показателю IoU регионов. Аналогично, для каждого из этих регионов формируется два набора векторов признаков одинакового размера, соответствующих фронтальному изображению и BEV проекции, которые объединяются в один вектор с помощью операции поэлементного усреднения. Полученные векторы поступают на входы трех полносвязных нейронных сетей,

которые вычисляют контуры объекта, ориентацию объекта в пространстве и класс объекта. Для фильтрации пересекающихся обнаружений используется метод подавления немаксимумов.

2.2. PointRCNN

Метод обнаружения PointRCNN [13] включает в себя два этапа: на первом этапе осуществляется генерация трехмерных предполагаемых регионов, на втором этапе происходит корректировка полученных регионов в канонической системе координат (canonical coordinate system) для получения окончательных результатов обнаружения. Структурная схема данного метода представлена на рисунке 2.

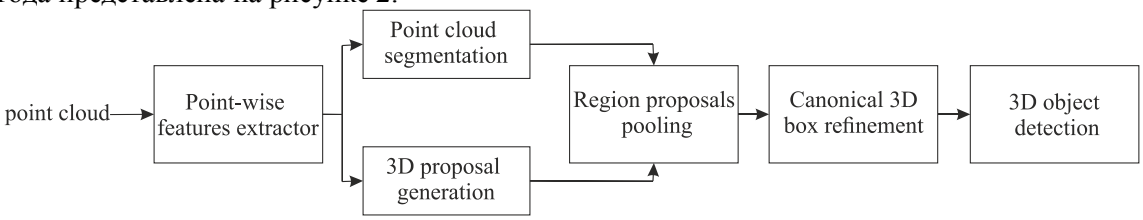

Рисунок 2. Структурная схема архитектуры метода PointRCNN.

Входными данными для данного метода является облако точек, полученное с помощью LIDAR датчика. Для формирования глобальных векторов признаков из входных данных авторы используют нейронную сеть PointNet++ [15].

Генерация трехмерных предполагаемых регионов на первом этапе осуществляется путем сегментации облака точек всей сцены на два класса: объект (foreground) и фон (background). Для каждой точки класса «объект» особым образом формируется трехмерный предполагаемый регион. Затем осуществляется расширение предполагаемых регионов. Эта операция необходима для того, чтобы учесть особенности пространственного положения объектов.

На первом шаге второго этапа осуществляется преобразование координат каждой точки предполагаемого региона к канонической системе координат. Затем для каждого региона и соответствующих ему точек формируется набор локальных векторов признаков. Локальные вектора признаков каждой точки полученных регионов и соответствующие им глобальные векторы признаков объединяются и используются непосредственно для получения трехмерных контуров объектов.

2.3. SECOND

Структурная схема метода SECOND (Sparsely Embedded CONvolutional Detection) [7] представлена на рисунке 3. Входными данными для данного метода является облако точек, полученное с помощью LIDAR датчика. Особенностью данного метода, является способ представления исходных данных. Для извлечения признаков из облака точек пространство точек разбивается на заданное число вокселей. В рамках данного метода воксель определяется как прямоугольный параллелепипед, фиксированного размера. Векторное представление вокселей формируется с помощью VFE (voxel feature encoding) [6] слоя нейронной сети.

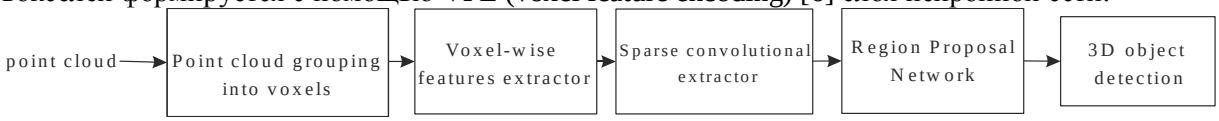

Рисунок 3. Структурная схема архитектуры метода SECOND.

Полученные векторные описания поступают на вход разряженной сверточной нейронной сети (Sparse Convolutional Network), которая используется для формирования двумерной карты признаков, учитывающей данные соответствующие оси z облака точек.

Полученные карты признаков поступают на вход RPN сети SSD (single shot multibox detector) [16], выходные значения которой подаются на вход трех сверточных слоев, для получения класса объекта, его контуров и ориентации в пространстве.

3. Экспериментальные данные

В рамках данной работы исследование эффективности методов обнаружения трехмерных объектов проводилось как на реальных данных, так и на синтетических. В качестве объектов детектирования выступали автомобили.

Реальные данные были представлены широко известной базой KITTI [12], которая позволяет проводить исследования различных методов и алгоритмов компьютерного зрения, использующих данные полученные с монокулярной камеры, стереокамеры, датчиков LIDAR. База KITTI содержит 7481 экземпляров размеченных данных, которые были разделены на две части: обучающую выборку (3712 экземпляров) и тестовую (3769 экземпляров).

В качестве источника синтетических данных был использован симулятор с открытым исходным кодом CARLA (Car Learning to Act) [11]. Данный симулятор позволяет эмулировать различные дорожные ситуации для обучения систем управления беспилотными автомобилями. Беспилотный автомобиль в среде симуляции CARLA может быть оснащен различными датчиками, которые применяются на реальных автомобилях. В рамках данной работы были использованы следующие датчики: монокулярная камера и LIDAR датчик. CARLA позволяет проводить настройку множества параметров для каждого датчика. Данные параметры были заданы в соответствии с параметрами, используемыми в базе KITTI. Было сгенерировано 7900 экземпляров синтетических данных, которые аналогичным образом были разделены на две выборки.

4. Экспериментальные исследования

При проведении экспериментальных исследований параметры каждого из рассматриваемых метода были заданны с учетом рекомендаций их авторов.

Для оценки качества прогноза использовалась средняя точность (average precision) обнаружения, при граничном значении метрики IoU равным 0.7. Метрика IoU вычисляется следующим образом:

$$IoU\ (p, t) = \frac{\left| p \cap t \right|}{\left| p \cup t \right|},$$

где p – это минимальный ограничивающий прямоугольный параллелепипед объекта, полученный с помощью метода детектирования, t – это истинный минимальный ограничивающий прямоугольный параллелепипед объекта.

Авторы базы данных KITTI ввели классификацию распознаваемых объектов по сложности распознавания (easy, moderate, hard), которая учитывает размер автомобиля на изображении и соотношение между его видимой части и не видимой. Экспериментальные исследования были проведены на сложности «easy».

На первом этапе проводимых экспериментальных исследований была исследована эффективность методов обнаружения трехмерных объектов на реальных данных. В качестве обучающей и тестовой выборки были использованы данные базы KITTI. Результаты представлены в таблице 1.

Исходя из полученных результатов, можно сделать вывод, что каждый из анализируемых методов работоспособен и демонстрирует высокое качество обнаружения. Лучший результат был показан методом PointRCNN.

Таблица 1. Сравнение методов детектирования трехмерных объектов на реальных данных.

Метод	Средняя точность детектирования на двумерном изображении	Средняя точность детектирования на BEV проекции	Средняя точность детектирования трехмерного объекта
AVOD	89.77	89.23	76.87
PointRCNN	**98.63**	**90.19**	**88.89**
SECOND	97.06	89.87	88.01

На следующем этапе экспериментальных исследований метода был проведен анализ эффективности рассматриваемых методов на синтетических данных, полученных из симулятора CARLA. Результаты представлены в таблице 2.

Таблица 2. Сравнение методов детектирования трехмерных объектов на синтетических данных.

Метод	Средняя точность детектирования на двумерном изображении	Средняя точность детектирования на BEV проекции	Средняя точность детектирования трехмерного объекта
AVOD	**80.97**	**81.17**	**78.52**
PointRCNN	17.78	1.00	1.64
SECOND	0.34	0.65	0.65

Исходя из полученных результатов, методы, использующие только данные LIDAR датчика, абсолютно не применимы на синтетических данных симулятора CARLA. В то же время метод AVOD, который также использует данные, полученные с помощью монокулярной камеры, демонстрирует хороший результат. Пример результата работы данного метода представлен на рисунке 4. Такая колоссальная разница в качестве работы методов PointRCNN и SECOND на реальных и синтетических данных может объясняться тем, что характеристики облака точек, полученные датчиком LIDAR в среде симуляции CARLA, существенно отличаются от соответствующих им характеристик облака точек реальных данных.

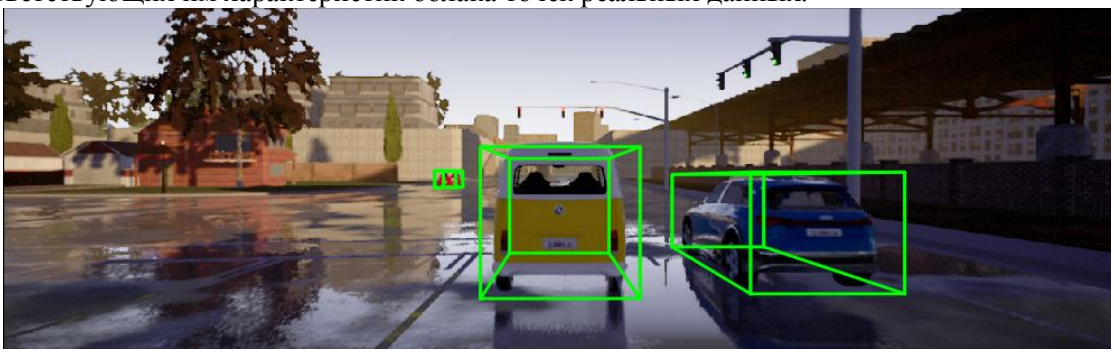

Рисунок 4. Пример результата работы метода AVOD на синтетических данных.

Так же было проведено исследование качества работы данных методов на реальных данных при их обучении на синтетических данных. Полученные результаты представлены в таблице 3.

Таблица 3. Сравнение методов детектирования трехмерных объектов на реальных данных при их обучении на синтетических данных.

Метод	Средняя точность детектирования на двумерном изображении	Средняя точность детектирования на BEV проекции	Средняя точность детектирования трехмерного объекта
AVOD	**17.24**	11.03	**4.59**
PointRCNN	0.16	0.06	0.06
SECOND	10.64	**15.25**	0.76

Исходя из полученных результатов, можно сделать вывод о том, что обучение данных методов только на синтетических данных не позволяет получить приемлемое качество детектирования на реальных данных. Даже метод AVOD, показавший хорошие результаты работы отдельно на реальных и синтетических данных, в данном случае оказался

неработоспособен. Таким образом, обучение рассматриваемых методов только на синтетических данных не позволяет их применить на реальных данных.

5. Заключение

В работе рассмотрены известные методы обнаружения трехмерных объектов. Проведены экспериментальные исследования их эффективности на реальных и синтетических данных. Представлены выводы о возможности применения рассмотренных методов детектирования трехмерных объектов, обученных на синтетических данных, на реальных данных.

Дальнейшие исследования могут быть направлены на анализ методов генерации реалистичных исходных данных для методов детектирования трехмерных объектов из синтетических данных с использованием генеративных нейронных сетей.

6. Благодарности

Работа выполнена при частичной финансовой поддержке грантов РФФИ № 18-29-03135-мк, № 18-07-00605 А.

7. Литература

[1] Paden, B. A survey of motion planning and control techniques for self-driving urban vehicles / B. Paden, M. Cap, S.Z. Yong, D. Yershov, E. Frazzoli // IEEE Transactions on intelligent vehicles – 2006. – Vol. 1. – P. 33-55.

[2] Mottaghi, R. A coarse-to-fine model for 3D pose estimation and sub-category recognition / R. Mottaghi, Y. Xiang, S. Savarese // Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) – 2015. – Vol. 1. – P. 418-426. DOI: 10.1109/CVPR.2015.7298639.

[3] Chabot, F. Deep MANTA: A Coarse-to-Fine Many-Task Network for Joint 2D and 3D Vehicle Analysis from Monocular Image / F. Chabot, M. Chaouch, J. Rabarisoa, C. Teuliere, T. Chateau // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Vol. 1. – P. 1827-1836. DOI: 10.1109/CVPR.2017.198.

[4] Li, B. GS3D: An Efficient 3D Object Detection Framework for Autonomous Driving / B. Li, W. Ouyang, L. Sheng, X. Zeng, X. Wang // Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – Vol. 29(4). – P. 1019-1028. DOI: 10.1109/CVPR.2019.00111.

[5] Mousavian, A. 3D Bounding Box Estimation Using Deep Learning and Geometry / A. Mousavian, D. Anguelov, J. Flynn, J. Kosecka // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Vol. 1. – P. 5632-5640. DOI: 10.1109/CVPR.2017.597.

[6] Zhou, Y. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection / Y. Zhou, O. Tuzel // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2018. – Vol. 1. – P. 4490-4499. DOI: 10.1109/CVPR.2018.00472.

[7] Yan, Y. SECOND: Sparsely Embedded Convolutional Detection / Y. Yan, Y. Mao, B. Li // Sensors. – 2018. – Vol. 18(10). – P. 3337. DOI: 10.3390/s18103337.

[8] Yang, B. PIXOR: Real-time 3D Object Detection from Point Clouds / B. Yang, W. Luo, R. Urtasun // Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2018. – Vol. 1. – P. 7652-7660. DOI: 10.1109/CVPR.2018.00798.

[9] Chen, X. Multi-view 3D Object Detection Network for Autonomous Driving / X. Chen, H. Ma, J. Wan, B. Li, T. Xia // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – Vol. 1. – P. 6526-6534. DOI: 10.1109/CVPR.2017.691.

[10] Joint 3d proposal generation and object detection from view aggregation [Electronic resource]. – Access mode: https://arxiv.org/pdf/1712.02294 (16.12.2019).

[11] Dosovitskiy, A. CARLA: An Open Urban Driving Simulator / A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez V. Koltun // Proceedings of Machine Learning Research. – 2017. – Vol. 78 – P. 1-16.

[12] Geiger, A. Are we ready for autonomous driving? The KITTI vision benchmark suite / A. Geiger, P. Lenz, R. Urtasun // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2012. – Vol. 1. – P. 3354-3361. DOI: 10.1109/CVPR. 2012.6248074.
[13] Shi, S. PointRCNN: 3D Object Proposal Generation and Detection From Point Cloud / S. Shi, X. Wang, H. Li // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – Vol. 1. – P. 770-779. DOI: 10.1109/CVPR.2019.00086.
[14] Very Deep Convolutional Networks for Large-Scale Image Recognition [Electronic resource]. – Access mode: https://arxiv.org/pdf/1409.1556 (16.12.2019).
[15] PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space [Electronic resource]. – Access mode: https://arxiv.org/pdf/1706.02413 (16.12.2019).
[16] Liu, W. SSD: Single Shot MultiBox Detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg // Computer Vision – ECCV. – 2016. – Vol. 1. – P. 21-37. DOI: 10.1007/978-3-319-46448-0_2.

A comparison of 3D objects detection methods for an autonomous car driving problem

A.A. Agafonov[1], A.S. Yumaganov[1]

[1]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. The task of autonomous vehicles control is one of the most immediate tasks both from a research and a practical point of view. To solve it, it is necessary to consider a number of subtasks, including localization and mapping, detection of static and dynamic objects, planning of maneuvers, control, etc. This paper discusses one of the most important tasks of autonomous vehicle control system: the task of detecting and localizing three-dimensional objects. All existing detection methods use information from various vehicle sensors: cameras and lidars. The results of experimental studies of these methods on real and synthetic data are presented. Synthetic data was obtained using CARLA – software for modelling and research of autonomous driving.